Claims

1. Process for the determination of interacting biomolecules characterized in that similar patterns of variation between two or more positions of at least two biomolecules are used.

2. Process for the determination of interacting biomolecules, characterized in that

a) a first group is provided comprising sequences representing homologous biomolecules,

b) at least one second group is provided comprising sequences representing homologous biomolecules,

c) group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and

d) the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

3. Process according to claim 2, characterized in that the probability of the interaction is calculated as predicted interaction value.

4. Process according to claim 2 or 3, characterized in that the interacting biomolecules are those with a positive predicted interaction value.

SUB A1

5. Process according to any of claims 2 to 4, characterized in that any of the second group(s) is converted into the first group and the first group is converted into a second group and group correlation values between the sequences of this new first group and

the sequences of any of the second group(s) which also comprises the former first group, are determined.

6. Process according to any of claims 2 to 5, characterized in that
site correlation values within each of the sequences within the first group and/or site correlation values within each of the sequences within the second group(s) are determined and said site correlation values are used for the calculation of the probability of interaction and/or for the calculation of the predicted interaction value of the sequence represented biomolecules.

7. Process according to claim 6, characterized in that the site correlation values are correlation values for substitutions within the sequences

8. Process according to any of claims 2 to 7, characterized in that

SUB A2

each sequence of each of said groups is fused to each other to form fused sequences comprising at least one sequence of the first group and at least one sequence of any second group(s),

the correlation values within these fused sequences are determined, and

the correlation values are used as group correlation values for determining the predicted interaction value and/or the probability of interaction.

9. Process according to any of claims 2 to 8 characterized in that
correlation values are determined by

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between equivalent positions of their position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

10. Process according to claim 9, characterized in that the standard distances matrix is the scoring matrix by McLachlan.

11. Method for the determination of interacting biomolecules which comprises processing data of at least a first set of data and at least a second set of data to output data

wherein each of the sets of data comprises independently and individually at least one or more elements,

wherein each of the elements represents the sequence of a biomolecule,

wherein the elements of a single set of data represent a group of homologous biomolecules,

wherein the output data comprises at least one pair of elements with one part of the pair of elements comprising at least one element from the first set of data and the other part of the pair of elements comprising at least one element from the second set of data,

characterised in that

- a group correlation values data set is created comprising group correlation values which are determined between the sequences of the first set of data and at least the second set of data;

- an interaction probability data set is created by retrieving group correlation values from the group correlation values data set and determining the probability of interaction of the biomolecules based on the group correlation values; and

at least some of the elements from the first and at least the second set of data which have been used to create the group correlation values and the interaction probability therefrom form the output data.

12. Method according to claim 11, characterized in that the probability of the interaction is calculated as predicted interaction value.

13. Method according to claim 11 or 12, characterized in that the elements the predicted interaction value of which is positive, are interacting biomolecules.

14. Method according to any of claims 11 to 13, characterized in that

any of second set(s) of data is converted into the first set of data and the first set of data is converted into a second set of data, and

group correlation values are determined between the sequences of this new first set of data and the sequences of any of the second set(s).

15. Method according to any of claims 11 to 14, characterized in that

site correlation values within each of the sequences within the first set of data and/or site correlation values within each of the sequences within the second set(s) of data are determined, and

said site correlation values form a set-specific site correlation value data set.

16. Method according to claim 15, characterized in that the set-specific site correlation value data set is used to calculate the probability of interaction of and/or to calculate the predicted interaction value of the sequence represented biomolecules.

17. Method according to claim 15 or 16, characterized in that the site correlation values are correlation values for substitutions within the sequences.

18. Method according to any of claims 11 to 17, characterized in that

a fused element set of data is generated by combining each element of the first set of data individually with each element of any of the second set(s) of data, and

attributing each fused element individually to the fused element set of data.

19. Method according to claim 18, characterized in that

the correlation values are determined within the various positions of a single element of the fused element set of data, and

the correlation values are used as group correlation values for determining the probability of the interaction of and/or predicted interaction value(s) of the biomolecules.

20. Method according to any one of claims 11 to 19, characterized in that the correlation values are determined by

SUB A5

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between equivalent positions of their position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

21. Method according to claim 20, characterized in that the standard distances matrix is the scoring matrix by McLachlan.

22. Method according to any of claims 11 to 21, characterized in that the first set of data and/or second the second set(s) of data are retrieved from a medium which is selected from the group comprising databanks, linked databanks, textual data and sets of data generated by an analytical instrument.

SUB A6

23. Method according to any of claims 11 to 22, characterized in that the set(s) of data comprise alligned sequences.

24. Method according to any of claims 11 to 23, characterized in that the output data are output control characters for a target medium.

25. Method or process according to any of claims 2 to 24, characterized in that the sequences of the first group or second group(s) or first set of data or second set(s) of data are selected from the group comprising DNA sequences, RNA sequences and amino acid sequences.

26. Method or process according to any of claims 2 to 25, characterized in that the number of sequences comprised in any of the groups or any of the sets of data is at least , preferably at least 11.

27. Method or process according to any of claims 2 to 26, characterized in that the sequences are homologous sequences.

28. Method or process according to claim 27, characterized in that the homologous sequences stem from different origins.

29. Method or process according to claim 27, characterized in that the homologous sequences in the first set of data and in the second set of data stem from the same origin and/or the homologous sequence in the first group and in the second group stem from the same origin.

30. Method or process according to any of claims 27 to 29, characterized in that the homologous sequences are homologous genes.

Sub A7

31. Method or process according to claim 30, characterized in that the homologous genes are orthologs.

32. Use of the method according to any of claims 11 to 31 for the simulation of biomolecule interaction.

33. Use according to claim 32 wherein the interacting biomolecules are those with a positive predicted interaction value determined by a process or method according to any of the preceeding claims.

34. Pairs of interacting biomolecules determined according to a method or process according to any of the claims 2 to 31.

35. Data structure readable by a computer, said data structure being generated by a process or a method according to any of claims 2 to 31.

36. Computer readable medium for embodying or storing therein data readable by a computer, said medium comprising one or more of the following:

a data structure generated by executing a process or a method according to any of claims 2 to 31;

Computer program code means which is adapted to cause a computer to execute a process or method according to any one of claims 2 to 31.

37. Computer program product comprising the computer readable medium according to claim 36.

38. Database containing information on interacting sequence pairs generated by applying the process or method according to any of the claims 2 to 31.

39. Database according to claim 38, wherein the database is an organism/species specific database.

40. Computer system comprising an execution environment for running the process or method according to any of the claims 2 to 31.

41. Device for simulating the interaction of biomolecules represented by their sequences which comprises

a loading device for making available the sets of data according to any of the claims 11 to 31,

a processing device for performing the method according to any of the claims 11 to 31,

an output device for receiving the output data generated by the processing device.